PUBLIC HEALTH REPORTS

VOL. 36. FEBRUARY 18, 1921 No. 7

INFLUENZA STUDIES.

By RAYMOND PEARL, Ph. D., D. Sc., LL. D., Professor of Biometry and Vital Statistics, School of Hygiene and Public Health, Johns Hopkins University; Consultant in Vital Statistics and Epidemiology, United States Public Health Service.

II. FURTHER DATA ON THE CORRELATION OF EXPLOSIVENESS OF OUTBREAK OF THE 1918 EPIDEMIC.[1]

I. Introduction.

In the first of these Studies [2] it was shown that there was a definite and sensible net correlation between explosiveness of outbreak of the epidemic, as measured by an epidemicity index, and the normal death rate from certain organic and chronic diseases. Because of the importance of the subject it has been thought desirable to reexamine critically the data, making use of more refined quantitative measures of the several variables dealt with. It is the object of the present paper to give the results of this re-study of the problem. As before, the basic data are from the large American cities for which weekly data were furnished during the epidemic, by the Bureau of the Census.[3] There is now in progress in this laboratory an extended study of the same problems on the basis of data from the 96 great towns of England and Wales, as well as further studies on the American data.

Before taking up the detailed matters of the present study, I should like to call attention briefly to some methodological considerations which lie at the foundation of this and other papers in this series. It is hoped that in this way the nonmathematical reader may gain a more adequate conception of the real meaning of the results.

The most useful general method of acquiring knowledge of dynamic phenomena is unquestionably the experimental method. When we deal with phenomena of human biology, there is a wide range of matters in which the laboratory experimental method is, in the nature of the case, ruled out. Unfortunately, one can not breed homozygous strains of men at will for experimental purposes, nor subject them

[1] Papers from the Department of Biometry and Vital Statistics, School of Hygiene and Public Health, Johns Hopkins University, No. 21.

[2] Pearl, R., Influenza Studies. I. On Certain General Statistical Aspects of the 1918 Epidemic in American Cities: Public Health Reports, vol. 34, No. 32, pp. 1743-1783, 1919. Reprint No. 548, pp. 1-43. All citations in this paper are in terms of the pagination of the reprint.

[3] It is a pleasure to express my indebtedness, for help with the extremely laborious computation of multiple correlation coefficients used in this work, to my assistants Mr. John Rice Miner, Mrs. Charmian Howell, and Miss Agnes L. Bacon.

methodically to desired environmental conditions. In studying most problems of human biology resort must be had to some form of the statistical method. This is fundamentally a descriptive method, and hence in many of its phases is ill adapted to the analysis of dynamically active events. There is, however, one branch of the statistical calculus which offers certain methodological possibilities which I think have been generally overlooked. Let us consider this.

The essence of the experimental method, as practiced in the laboratory, and in theory, is that, of the multitude of variables conditioning a phenomenon, as many as possible are, by appropriate methods, held constant while one variable (or at most a very few selected variables) is allowed to vary and the results are noted. One may then deduce the relative significance of the selected variable in determining the phenomenon under observation. Now we frequently hear in scientific discussions about the experiments that Nature makes. Actually the true conditions of an experiment are rarely if ever realized in the course of natural events. It is just because Nature permits manifold and haphazard changes in *all* variables at the same time that recourse must be had to the method of experimental control in the laboratory. What is needed in order to interpret the results, in the experimental sense, and determine the meaning of the manifold and ceaseless changes and variations in the flow of naturally determined events, is some method of picking out of the manifold some selected *constant* conditions of a series of variables, and then measuring the extent and character of the variations in a *single* selected variable, whose true relative influence upon the phenomenon it is desired to know, while all these other variables are held constant. If this can be done we shall have realized all the epistemological advantages of the experimental method as practiced in the laboratory, and have freed ourselves at the same time from the limitations which in so many cases inhere in the material itself and make the laboratory type of experimental inquiry impossible. In other words, we shall have let Nature perform the experiment, in the sense of determining the phenomena, in her own way, while we evaluate the results in critically analytical terms of precisely the same sort and meaning as those in which we evaluate the results of a laboratory experiment.

Now exactly this epistemological boon is actually afforded in the method of partial or net correlation, if properly handled. This analytically simple, if geometrically complex, calculus enables one, out of a manifold complex of variables operating in an entirely uncontrolled and natural manner, to determine the variation of any selected single variable, or the correlation of any selected pair for *constant* conditions or values of the other variables in the complex. I judge that the epistemological possibilities of this method are not

yet fully grasped by scientific men generally. When they are I believe it will rank as a fundamental method of acquiring knowledge, combining certain of the advantages of the descriptive or historical and of the experimental methods. It seems to me much more effectively to justify Royce's [4] eulogy of the statistical method than any of the arguments which he advanced. The best elementary, but at the same time adequate, account of the method is that of Yule.[5]

The problem with which this paper has to do may be stated in this way: It is an obvious fact that the large American cities varied enormously among themselves in respect to the explosiveness of outbreak of epidemic mortality in the autumn of 1918. What factors, environmental or other, were significant in determining or influencing this variation? Or, put in another way, what factors of the numerous other respects in which these 34 large cities differ from one another can be shown to be significantly correlated with the observed differences in explosiveness of outbreak of epidemic mortality?

II. Variables Discussed.

The variable phenomena or attributes discussed in the present paper are listed below, together with the subscript numbers by which they will be designated in this and subsequent papers in this series.

Subscript No.	Variable.
1.	Explosiveness of outbreak of epidemic mortality as measured by an epidemicity index I_6.
3a.	Normal death rate from pulmonary tuberculosis.
3b.	Normal death rate from organic diseases of the heart.
3c.	Normal death rate from acute nephritis and Bright's disease.
3d.	Normal death rate from typhoid fever.
3e.	Normal death rate from cancer and other malignant tumors.
3f.	Normal death rate from all causes.
4.	Age distribution of population.
5.	Sex ratio of population.
6.	Density of population.
7.	Latitude.
8.	Longitude.
9.	Rate of growth of population, 1900-1910.

In the following paragraphs these variables are defined and discussed in detail.

1. In the first of these Studies a number of indices of explosiveness of outbreak of epidemic mortality were considered and the one finally adopted, I_5, was called the peak-time ratio and defined as follows:

$$I_5 = \frac{P - M}{T},$$

[4] Royce, J., The Mechanical, the Historical, and the Statistical: Science, N. S. 1914.
[5] Yule, G. Udny, An Introduction to the Theory of Statistics, Fifth Edition, pp. 228–253, London, 1919.

where P denotes the maximum peak mortality rate observed during the duration, T, of the epidemic, which was defined as follows: "The epidemic mortality was considered to have begun in any city on the date when the mortality curve for that city first passed outside the range of fluctuation exhibited by the curve between the week ending July 6, 1918, and the end of the week immediately preceding the epidemic rise of the curve. The mortality of the first epidemic outbreak was considered to have ended on the date when the curve again passed within the same range of fluctuation." M' is the mean weekly annual death rate in the period from July 6, 1918, to the outbreak of the epidemic.

This index, I_5, increases as the explosiveness of the outbreak increases. After the publication of the paper, however, it was pointed out by Dr. W. H. Frost and Mr. Edgar Sydenstricker, of the United States Public Health Service, that it was open to some criticism as a measure of explosiveness of epidemic outbreak, in the strictest sense of the term. The point of criticism was that inasmuch as T included the whole time within which the mortality curve was outside the normal range, the value of the index would be influenced by both the ascending and descending limbs of the epidemic curve; whereas if it is strict explosiveness of *outbreak* that we wish to measure, only the ascending limb is of moment. Reflection shows that the point is well taken, and consequently in the present study we have used an index I_6, for which the symbolic expression is:

$$I_6 = \frac{P - M'}{T'}.$$

Here the letters have the same significance as before except that T' is the number of weeks elapsed between (a) the date when the mortality curve first passed outside the range of fluctuation exhibited by the curve between the week ended July 6, 1918, and the end of the week immediately preceding the epidemic rise of the curve, and (b) the week in which the mortality curve attained its first epidemic peak. In other words, there is now included in the epidemicity index only the ascending limb of the epidemic curve. As a matter of fact, in the American cities here dealt with, little practical difference is made in any conclusions regarding the epidemic whether one uses I_5 or I_6, but there can be no question that theoretically I_6 is the superior value, and consequently we have used it in this part and the following parts of these Studies.

2. The subscript 2 refers to total destructiveness of the epidemic, a variable not discussed in the present paper, but defined in the next following of these Studies.

3. The subscript 3 refers to a normal death rate in the community from one of the causes specified by letters. In the earlier study the

death rate of the community for a single year was taken as indicative of the normal for the purposes of that work. In the present study we have taken instead the mean annual death rate from each of the specified causes for the three years 1915, 1916, and 1917. It is beyond question that the three-year average will give a much more accurate representation of the prevailing normal rate of mortality from each of these diseases in the community just preceding the epidemic than will the rate for any single year.

4. Age distribution of the population. In the first of these Studies there was used, as a single numerical index of the age distribution, a quantity which measured the extent to which each city deviates in the age constitution of its population from a fixed standard, but did not tell the nature or kind of deviation. Since the publication of that paper there has been devised a new and more adequate index [6] of the age constitution of the population. We have adopted as an age-constitution index the function

$$\phi = S\left[\frac{\Delta^2}{P}\right](M - M_p)$$

where Δ is the deviation for each of six age groups (viz, 0–4, 5–14, 15–24, 25–44, 45–64, 65 and over) of the percentage of the actual population of each city in 1910 in each age group, from the percentage in the same group in the standard population of Glover's [7] life table, denoted in the formula by P; S denotes summation of all six values; M = mean age of living population in any community; M_p = mean age of persons in a stationary population unaffected by migration and which, assuming the mortality rates of Glover's life table, would result if 100,000 persons were born alive uniformly throughout each year (M_p calculated from L_x line of Glover's table (p. 16), = 33.796 years).

This procedure simply multiplies our former index χ^2 by the difference (given its proper sign) between the mean age of the observed population and the mean age of the standard population on the basis of which χ^2 was calculated. Since, in fact, the mean age of any actual urban population is never likely to be as great as the mean age of the stationary population chosen as a standard of reference, the actual values of ϕ will practically always be negative for cities. The smaller these negative values are numerically, the greater will be the proportion of older persons in the population concerned. In short, this function ϕ tells us not only the degree to which a given population deviates in its age distribution from a fixed standard age distribution, but also the nature of this deviation, whether on the one hand in the direction of a relative excess of

[6] Cf. Pearl, R., On a Single Numerical Index of the Age Distribution of a Population: Proc. Nat. Acad. Sci., vol. 6, pp. 427–431, 1920.

[7] Glover, J. W., United States Life Tables, 1910, Bureau of the Census, 1916.

aged, or on the other hand in the direction of a relative excess of the young. Theoretically it is possible for two populations differing from one another in a compensatory way to give the same values for the index ϕ. But two populations which differ in age distribution in any fundamental respect which could affect appreciably crude death rates will, in all populations I have been able to test, give different values of ϕ, provided the age classification from which the function is calculated is finely enough divided.

5. As an index of the sex distribution of the population, the male sex ratio was expressed as the ratio of males to 100 females in 1910.

6. The density of population in each city was calculated from data furnished in the "Financial Statistics of Cities," issued annually by the Bureau of the Census, and was expressed as the number of persons per acre of land area within the legally defined limits of the city.

7 and 8. In the earlier study the geographical position of the cities was indicated by the linear distance of each from Boston, as measured on the map. This was recognized as a very rough approximation. The interest in having some measure, in a study of this kind, of geographical position is twofold: First, that which arises from purely epidemiological considerations, namely, as affording, in relation to time, an index of the rate of spread of an infectious epidemic disease from a primary focus; and secondly, the fact that geographical position, especially latitude, is a rough but on the whole fairly accurate index of general climatological conditions. It was decided to make the expression of geographical position more accurate in the present study, and consequently there were included as definite variables the latitude and longitude of each of the cities considered.

9. Rate of growth of population 1900–1910. The reason for the inclusion of this variable in the study was twofold. Primarily this may be taken as a rough but probably fairly accurate index of the degree of industrialization of a city. In general, those cities which are growing most rapidly in population are those in which the most rapid industrial development is taking place. It would be much better if the rate of growth of the population between 1910 and the outbreak of the epidemic could have been used; but accurate data are lacking, nor will they be available until the results of the 1920 census are published. Consequently this variable must be regarded as a rough approximation to one that we should like to measure more accurately, namely, the present state and recent rate of industrial development. In the second place, rate of growth is a definite biological characteristic of a population,[8] and as such

[8] *Cf.* Pearl, R., and Reed, L. J., On the Rate of Growth of the Population of the United States since 1790 and its Mathematical Representation: Proc. Nat. Acad. Sci., vol. 6, pp. 275–288, 1920.

worthy of inclusion in any study relating an epidemic disease to demographic conditions.

III. Data.

The actual data used in the correlations are given in Table 1. Since in other work it was desired to correlate destructiveness of the epidemic, as measured by the 25-week excess mortality, with the other variables, only 34 cities could be used, because only for that number are the excess mortality figures available.

TABLE 1.—*Data for correlation of characteristics of cities with explosiveness of epidemic influenza mortality.*

(The subscript numbers at the heads of columns correspond to the list of variables given on p. 275.)

City.	1.	3a.	3b.	3c.	3d.	3e.	3f.	4.	5.	6.	7.	8.	9.
Albany, N. Y.	32.2	224.5	236.4	180.1	10.8	137.6	1,947.0	−10.73	92.9	8.89	42.65°	73.75°	6.5
Atlanta, Ga.	5.3	113.9	121.6	159.2	19.5	66.5	1,551.2	−82.71	92.7	11.42	33.73	84.33	72.3
Baltimore, Md.	43.4	202.7	195.9	179.0	18.1	107.5	1,810.3	−31.80	92.4	30.57	39.28	76.62	9.7
Boston, Mass.	28.8	143.8	211.5	101.0	3.9	117.0	1,651.6	−31.05	96.7	27.36	42.36	71.06	19.6
Buffalo, N. Y.	21.1	145.0	155.9	125.0	10.3	97.8	1,593.4	−48.61	100.6	18.97	42.88	78.92	20.2
Cambridge, Mass.	17.9	178.8	183.9	73.6	2.7	118.8	1,354.4	−30.26	91.7	28.23	42.38	71.13	14.1
Chicago, Ill.	13.2	139.5	159.1	110.4	4.1	88.8	1,457.1	−68.73	105.3	20.28	41.88	87.60	28.7
Cincinnati, Ohio.	9.2	206.8	202.5	162.0	4.8	111.5	1,618.0	−22.93	95.4	9.10	38.14	84.42	11.6
Cleveland, Ohio.	17.7	132.8	118.8	90.0	7.0	82.4	1,432.1	−74.51	103.6	20.08	41.50	81.70	46.9
Columbus, Ohio.	11.0	131.8	155.2	88.0	11.4	105.1	1,498.3	−29.08	101.5	15.18	40.00	83.00	44.6
Dayton, Ohio.	28.8	139.2	184.3	103.2	17.3	107.7	1,492.7	−24.10	101.9	12.65	39.73	84.18	36.6
Fall River, Mass.	27.8	139.6	164.0	108.5	14.3	83.9	1,639.4	−73.39	93.4	5.91	41.70	71.15	13.8
Grand Rapids, Mich.	2.5	76.3	145.4	94.7	19.0	98.4	1,258.8	−27.95	97.4	11.85	42.97	85.70	28.6
Indianapolis, Ind.	8.6	164.2	185.7	111.1	17.1	95.3	1,550.2	−24.23	98.7	10.96	39.67	86.13	38.1
Louisville, Ky.	21.5	160.1	164.7	160.1	14.2	83.5	1,542.1	−3?.19	94.1	16.61	38.20	85.70	9.4
Los Angeles, Calif.	10.0	182.1	153.0	103.8	4.7	103.5	1,238.2	−13.61	103.9	2.49	34.08	118.20	211.5
Lowell, Mass.	24.7	112.5	157.7	90.7	11.2	79.8	1,681.9	−34.91	94.1	13.63	42.65	71.32	11.9
Milwaukee, Wis.	4.9	82.4	100.7	73.7	8.8	85.3	1,214.4	−62.17	102.8	26.92	43.05	87.95	31.0
Minneapolis, Minn.	4.1	120.4	114.1	101.3	6.5	93.8	1,189.4	−55.39	100.2	11.27	44.97	93.30	48.7
Nashville, Tenn.	41.5	189.3	198.7	123.1	30.0	83.3	1,683.7	−48.32	89.6	10.11	36.15	86.80	36.5
Newark, N. J.	14.1	151.0	148.3	137.1	4.3	88.5	1,412.2	−62.54	99.6	27.52	40.75	74.17	41.2
New Haven, Conn.	11.1	103.6	185.4	125.1	13.0	112.1	1,650.0	−33.70	99.7	13.03	41.32	72.92	23.7
New Orleans, La.	34.1	270.9	224.4	245.6	22.3	95.5	1,983.0	−52.44	92.8	2.93	29.97	90.08	18.1
New York, N. Y.	12.8	160.5	165.3	128.3	4.6	85.0	1,384.7	−74.42	99.9	29.54	40.71	74.00	38.7
Oakland, Calif.	15.1	98.7	187.6	87.7	4.4	94.2	1,082.8	−17.18	108.7	6.41	37.75	122.33	124.3
Philadelphia, Pa.	47.9	169.9	203.5	178.3	6.9	97.6	1,630.1	−24.27	95.4	21.02	39.95	75.17	19.7
Pittsburgh, Pa.	21.3	115.9	134.5	94.3	10.4	90.7	1,694.0	−71.77	105.1	22.81	40.43	80.03	18.2
Providence, R. I.	14.0	133.3	152.8	130.2	6.4	99.1	1,527.0	−30.05	95.7	22.35	41.84	71.49	27.8
Rochester, N. Y.	13.9	95.7	203.1	138.0	4.7	109.8	1,456.8	−26.62	98.7	18.62	43.13	77.85	34.2
St. Louis, Mo.	5.4	132.2	141.3	171.9	8.1	97.3	1,458.0	−44.57	101.5	19.36	38.63	90.20	19.4
St. Paul, Minn.	4.5	103.6	118.9	73.7	5.0	86.2	1,093.0	−68.95	103.6	7.49	44.87	93.08	31.7
San Francisco, Calif.	25.4	170.0	257.2	132.5	5.7	138.0	1,546.5	−34.34	131.6	17.55	37.79	122.43	21.6
Toledo, Ohio.	17.9	173.7	185.2	88.7	18.0	103.9	1,705.1	−33.21	101.1	10.91	41.77	83.55	27.8
Washington, D. C.	30.7	185.5	235.1	176.2	12.7	110.5	1,795.6	−20.67	91.3	9.55	38.89	77.03	18.8

IV. Demographic and Environmental Correlation.

We come now to the consideration of the results. The net influence of the several demographic and environmental factors upon explosiveness of epidemic outbreak may be first considered. As we are obviously interested in getting at the *net* influence of each factor, such as age distribution of the population upon variation in the epidemicity index, while all the other factors for which we have data are held to constant values, we may pass at once to the fifth order correlation coefficients, without stopping for detailed consideration of the lower order coefficients leading to the final values. Such points as do need discussion in connection with these lower order coefficients will be brought out in connection with the fifth order results. The net

correlations between epidemicity index (subscript 1) and each of the six demographic and environmental factors taken one at a time, with the other five held constant, are exhibited in Table 2. It is to be understood in this, and all subsequent papers in this series, that subscript numbers to the right of the decimal point denote variables held constant, and subscript numbers to the left of the point denote the variables correlated. r is, of course, simply the conventional symbol of a correlation coefficient.

TABLE II.—*Net correlation of explosiveness of outbreak (I_6) with various demographic and environmental factors.*

Variable correlated with explosiveness (I_6).	r subscripts.	Coefficient.
Age distribution of population	14.56789	+0.281 ±0.107
Sex ratio of population	15.46789	−0.001 ±0.116
Density of population	16.45789	+0.099 ±0.115
Latitude of city	17.45689	−0.389 ±0.100
Longitude of city	18.45679	−0.085 ±0.115
Rate of growth of population, 1900–1910	19.45678	−0.288 ±0.106

Taking the several variables in order we note:

1. Explosiveness of outbreak of epidemic mortality can not be positively asserted to be significantly correlated with the age distribution of the population. The coefficient +0.281 is less than 3 times its probable error. The plus sign means, having regard to the method of calculating age indices explained above, that so far as there is any correlation, high values of the explosiveness index I_6 tended to be associated with populations having a higher proportion of *older* persons than the average. No one of the lower order correlations of explosiveness and age index had a value as much as 3 times its probable error. The highest coefficient in the series was the fourth order $r_{14\cdot 6789} = +0.300 \pm 0.105$. The zero order $r_{14} = +0.194 \pm 0.111$. All r_{14} correlations, whatever the secondary subscripts, within the group now under consideration, are positive. In the first of these studies (*loc. cit.*, p. 38) the explosiveness-age correlation of zero order, with the less exact age distribution index, was $r_{14} = -0.262 \pm 0.101$, again a value nearly but not quite certainly significant. With border line values such as these, one can only say that in cities constant in respect to sex ratio of population, density of population, position, and rate of recent growth of population, the age composition of the population may have a slight influence in determining explosiveness of outbreak of epidemic mortality, but at best the influence must be very small.

2. The sex ratio of the population is plainly not significantly correlated with epidemicity index. The fifth order coefficient has a value -0.001 ± 0.116, which is sensibly zero. This is an interesting example of how partial or net correlations may differ from total correla-

tions. The zero order coefficient between explosiveness (I_6) and sex ratio is $r_{15} = -0.307 \pm 0.105$. This is a probably significant value, but arises not from any direct relation of variables 1 and 5, but indirectly through the relation of both of these to the positional variables 7 and 8 (latitude and longitude). In the group of cities here dealt with there is a relatively high correlation between male sex ratio and longitude, and between sex ratio and latitude, when longitude is constant. These relations are shown in the following coefficients:

$r_{57} = +0.134 \pm 0.114$ (sex ratio and latitude).
$r_{58} = +0.678 \pm 0.062$ (sex ratio and longitude).
$r_{57 \cdot 8} = +0.607 \pm 0.073$ (sex ratio and latitude—longitude constant).
$r_{58 \cdot 7} = +0.808 \pm 0.040$ (sex ratio and longitude—latitude constant).

Because of the inverse and nearly equal correlations of epidemicity index with latitude and longitude, the sex-ratio correlation with epidemicity index is neutralized as soon as these other variables are brought into the system. Thus while we have $r_{15} = -0.307 \pm 0.105$, we get $r_{15 \cdot 78} = +0.023 \pm 0.116$, or practically zero. We may then safely conclude that the proportion of males (or of females) in the population of a city had no sensible direct influence in determining the explosiveness of outbreak of epidemic mortality.

3. The net fifth order correlation of explosiveness of outbreak of the epidemic mortality with density of population is again sensibly zero. This is true whatever the variables held constant within the group here discussed. The zero order coefficient is $r_{16} = +0.073 \pm 0.115$. Nowhere in the series does a coefficient having primary subscripts 16 rise to a value even approaching 3 times the probable error. This result, that density of population, which is the measure of urban crowding in this case, had nothing to do with determining the explosiveness of outbreak of the epidemic mortality, while surprising on grounds of purely *a priori* logic—always, by the way, most unsafe grounds—fully confirms with more critical data the result attained in the earlier study (*loc. cit.*, p. 36).

4. In the case of the correlation of explosiveness of outbreak of epidemic mortality with the latitude of the city, a very different result appears. The fifth order coefficient is $r_{17 \cdot 4568} = -0.369 \pm 0.100$. This is more than three times its probable error and is probably *statistically* significant; but before drawing any conclusions about its epidemiological significance we must critically look into its genesis. In the first place it must be noted that the 34 cities with which we are dealing are not scattered at random over the United States. All but three of them are either on or east of the Mississippi River. This distribution is well shown in the map exhibited as Figure 1.

Not only are the cities mostly in the eastern half of the country, but, what is more important, they nearly all fall in a fairly narrow northeast-southwest belt. How clearly this is so is shown by the

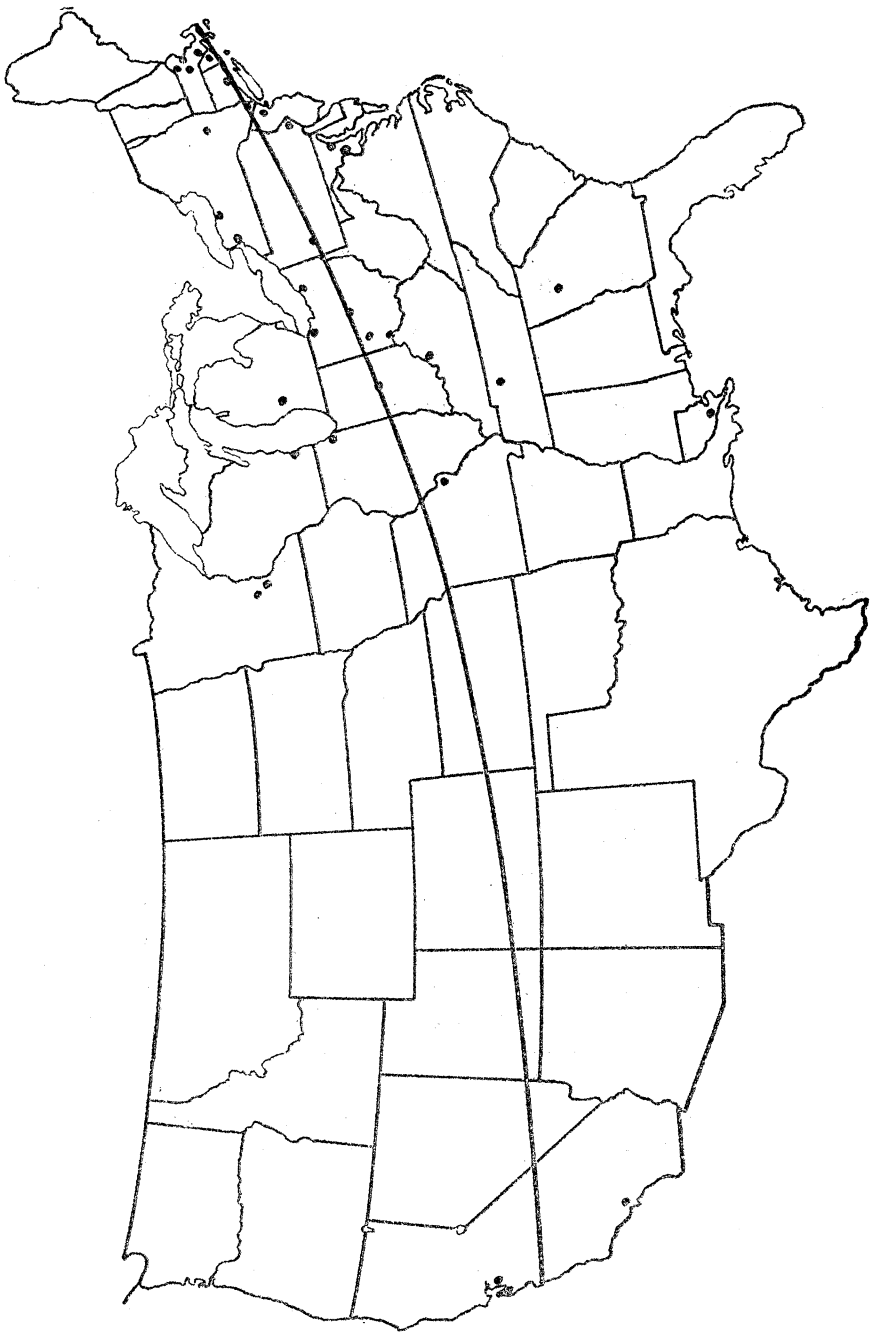

FIG. 1.—Outline map of the United States showing the location of the 34 cities used in this study.

regression of latitude on longitude, which is plotted on the map. The distribution of these cities is so far from random over the whole area of the country that there is a statistically significant correlation for these 34 cities, between latitude and longitude, the coefficient being $r_{78} = -0.404 \pm 0.097$. In words, what this means is that, within this group of 34 cities, in general, the farther north a city is the farther east it is.

Now, neither latitude nor longitude is alone significantly correlated with the epidemicity index I_6, as witness the zero order coefficients

$$r_{17} = -0.243 \pm 0.109$$
$$r_{18} = -0.229 \pm 0.110$$

But because of the relatively large value of r_{78} which has just been pointed out we get at once

$$r_{17.8} = -0.376 \pm 0.099 \text{ (significant)}$$
$$r_{18.7} = -0.369 \pm 0\ 100 \text{ (significant)}.$$

The same influence makes itself felt throughout the series of ascending order partial coefficients. Thus, the relevant second order coefficients are:

$r_{17.45} = -0.208 \pm 0.111$
$r_{17.46} = -0.303 \pm 0.105$
$r_{17.48} = -0.387 \pm 0.098$
$r_{17.49} = -0.407 \pm 0.097$
$r_{17.56} = -0.264 \pm 0.108$
$r_{17.58} = -0.319 \pm 0.104$
$r_{17.59} = -0.358 \pm 0.101$
$r_{17.68} = -0.381 \pm 0.099$
$r_{17.69} = -0.403 \pm 0.097$
$r_{17.89} = -0.421 \pm 0.095$

$r_{18.45} = -0.092 \pm 0.115$
$r_{18.46} = -0.242 \pm 0.109$
$r_{18.47} = -0.407 \pm 0.097$
$r_{18.49} = -0.074 \pm 0.115$
$r_{18.56} = +0.022 \pm 0.116$
$r_{18.57} = -0.245 \pm 0.109$
$r_{18.59} = +0.194 \pm 0.111$
$r_{18.67} = -0.333 \pm 0.103$
$r_{18.69} = -0.061 \pm 0.115$
$r_{18.79} = -0.160 \pm 0.113$

It is at once evident that most of the r_{17} or latitude coefficients are three or more time their probable errors. Most of the r_{18}, or longitude coefficients are, on the contrary, less than three times their probable errors, the only ones arising to a higher value being those carrying 7 as a secondary subscript and not all of those. In other words, we come here to a separating point between the correlations carrying 1 and 7 as primary subscripts and those carrying 1 and 8 in the same position. This divergence comes about from the different correlations of certain of the other variables with latitude and longitude. Thus we have, for sex-ratio correlations, $r_{75} = +0.134 \pm 0.114$ against $r_{85} = +0.678 \pm 0.062$. Age distribution index is not significantly correlated with either latitude or longitude in this group of cities. Density is significantly and about equally correlated with both, the coefficients being $r_{86} = -0.424 \pm 0.095$, and $r_{76} = -0.371 \pm 0.100$. Rate of growth in this group of cities is nearly twice as highly correlated with

longitude as with latitude, the signs of course being opposite. The coefficients are $r_{79} = -0.365 \pm 0.100$ and $r_{89} = +0.642 \pm 0.068$.

Without pursuing this complex trail further, certain things are clear. In the first place it is evident that it will be quite unsafe to draw biostatistical or demographic conclusions about, or on the basis of data from, the large American cities, without having critical regard for the fact here demonstrated that some of the most important of these characteristics are significantly correlated with the mere geographical position of the cities. In the second place, because of this fact, we can not be quite sure of the epidemiological significance of the fact that, in this group of cities, there is a statistically significant negative correlation between epidemiological index and latitude. Taken at its face value this coefficient means that, on the average, the outbreak of epidemic mortality was *more* suddenly explosive, the farther *south* the city, when the other factors of age distribution, sex ratio, density, and rate of growth of population were constant and equal. Whether this bespeaks a real and general biological phenomenon resting presumably upon a climatological base can not be critically determined until we can study the matter in a group of localities distributed in a more random manner than in the present sample, so that the correlations between latitude and longitude shall be more nearly zero in value. It is interesting to note, however, that the present result in respect to latitude correlation is in accord with general clinical and pathological experience. I am told by Dr. William H. Welch that it has long been recognized by clinicians and pathologists that acute respiratory infections, and particularly the pneumonias, tend to become more fatal as one passes from north to south. There is a splendid chance here for a critical statistical investigation of the matter, and with the rapid extension of the registration area into the South in recent years, adequate data will shortly be available.

5. Rate of growth of population of these cities in the decade 1900 to 1910 is a factor connected with age of city, with its industrial and commercial prosperity and activity, and indirectly with sanitation, because usually in cities growing very rapidly, sanitary arrangements tend to lag behind the need for them. The net fifth order correlation of this variable with explosiveness of outbreak of epidemic mortality is $r_{19.45678} = -0.288 \pm 0.106$. This is a border-line value, which can not be safely asserted to differ significantly from zero. Roughly speaking, a value as large as this would turn up purely by chance about 7 times in every 100 trials with samples of the size here dealt with. The gross, zero order coefficient between these variables is $r_{19} = -0.302 \pm 0.105$, again not a significant value. No one of the partial coefficients having 1 and 9 as primary sub-

scripts is certainly significant in comparison with its probable error, except such as carry also 7 as a secondary subscript. On the whole, it can be safely asserted that if the rate at which a city had recently been growing in population had anything at all to do with the degree of explosiveness of outbreak of the 1918 epidemic mortality, this influence must have been at the most extremely slight.

6. In general, the results of this more extended and critical study of the influence of these demographic and environmental factors confirm the earlier preliminary examination of the data. The only variable in the lot that has a statistically significant net correlation with the explosiveness index is latitude. In the former study, use was made of linear distance from Boston as a locality index instead of the more general and exact latitude and longitude plan here adopted. But the correlation result obtained with the cruder variables was the same in sense, and of about the same order of magnitude as the more refined net fifth order latitude correlations. I have already indicated fully the wisdom of caution for the present in drawing biological conclusions from this latitude correlation. It is possible that the age distribution of the population and the rate of recent growth of the city had some slight influence upon the explosiveness of outbreak of the epidemic mortality; but in either case the effect must have been so slight as to be negligible so far as any practical epidemiological significance is concerned.

V. Death Rate Correlations.

We may turn now to another series of correlations. In the first study it was shown that the most significant correlations of explosiveness of outbreak of epidemic mortality were with the normal death rates from certain primarily organic diseases. These results, because of their novelty and possibly far-reaching theoretical significance, have been thought particularly to need critical reexamination. As has already been pointed out, we have in the present study taken the mean death rate for three years as indicative of normal conditions instead of a single year as in the earlier work.

The sixth order coefficients are exhibited in Table III. These coefficients measure the net correlation existing between the epidemicity index I_6 and the specified normal death rate, when the cities involved have been made constant and equal in respect to age and sex constitution of population, to density and rate of recent growth of population, and to latitude and longitude.

TABLE III.—*Net correlation of explosiveness of outbreak (I_6) with the normal death rates from certain specified causes.*

Variable correlated with explosiveness (I_6): Death rate from—	r subscripts.	Coefficient.
All causes	13f. 456789	+0.572 ±0.078
Pulmonary tuberculosis	13a. 456789	+0.389 ±0.098
Organic diseases of the heart	13b. 456789	+0.562 ±0.079
Acute nephritis and Bright's disease	13c. 456789	+0.307 ±0.105
Typhoid fever	13d. 456789	+0.105 ±0.114
Cancer and other malignant tumors	13e. 456789	+0.141 ±0.113

It is at once evident that these correlations are of a generally different order of magnitude from those of Table II. Specifically we note:

1. The highest correlation is that for the death rate from all causes; but that for organic diseases of the heart is practically identical. These coefficients are more than seven times the probable error and are certainly significant. The corresponding zero order coefficients are:

$$\text{All causes,} \quad r_{13f} = +0.678 \pm 0.063$$
$$\text{Organic heart,} \quad r_{13b} = +0.642 \pm 0.068$$

From these values, in comparison with those of Table III, it appears that by making all six demographic and environmental factors constant, the correlation between explosiveness and the normal death rate from all causes or that from organic diseases of the heart, is not altered to a degree statistically significant. This comparison indicates the overwhelming importance of the biological factor which these death rates measure in determining the explosiveness of the outbreak of epidemic mortality, as compared with the demographic and environmental factors previously considered.

2. Next to organic diseases of the heart, pulmonary tuberculosis is the single cause having its normal death rate most highly correlated with explosiveness of outbreak of the epidemic. The net sixth order coefficient for this disease, however, is distinctly lower than that for organic diseases of the heart. Furthermore it is much more reduced by the process of making demographic and environmental factors constant, as is indicated by the fact that the zero order coefficient for this disease is $r_{13a} = +0.578 \pm 0.077$. The difference between this and the sixth order coefficient $r_{13a.456789} = +0.389 \pm 0.098$ is 0.189 ± 0.125. While not statistically significant in comparison with the probable error, it comes much nearer being so than the corresponding difference for the organic heart correlation, which is 0.080 ± 0.104.

3. In the case of the normal death rate from breakdown of the kidneys (acute nephritis and Bright's disease) the sixth order net coefficient is on the border line of probable statistical significance,

having a value just under three times its probable error. Again the process of making the six demographic and environmental factors constant has materially reduced the correlation between the normal death rate from these diseases of the kidneys and the explosiveness of outbreak of epidemic mortality. The zero order correlation here is $r_{13c} = +0.447 \pm 0.093$. The difference between the zero order and the sixth order coefficients is 0.140.

4. Turning to diseases of wholly different etiology from those dealt with up to this point, namely, typhoid fever and cancer, the correlations between the normal death rate from these diseases and the explosiveness of outbreak of epidemic mortality are found to be of an entirely different order of magnitude. In neither case, typhoid or cancer, is the coefficient sensibly different from zero, having regard to its probable error. Clearly, whatever factors, biological or environmental, or both, are measured by these death rates can not have had any sensible influence in determining the suddenness with which the mortality curve rose during the influenza epidemic. It is generally held on good grounds that the typhoid death rate is an excellent index of the general sanitary status of a community. If it may be so accepted in the present connection, the result just stated bears out in precise mathematical terms what was obvious to the thoughtful and candid observer at the time of the epidemic, namely, that the severity with which a city was hit by the epidemic bore no relation to its general sanitary status or to the efficiency of its health organization. In this connection there is a further interesting mathematical point regarding the typhoid correlation. The zero order gross correlation for this death rate with I_6 is $r_{13d} = +0.342 \pm 0.102$. This is a statistically significant correlation; but observe that as we make the cities constant in respect to demographic and environmental factors the reduction in the correlation is very great, ending with the sixth order coefficient at a value $r_{13d.456789} = +0.105 \pm 0.114$, a drop of 0.237 in the coefficient. But this is exactly what would be expected when it is recalled to what an extent the typhoid death rate of a community depends upon the environmental conditions in that community.

5. The normal cancer death rate is not significantly correlated with the epidemicity at any stage, nor is the correlation altered to any extent by making the demographic and environmental factors constant. This is indicated by the fact that the zero order coefficient is $r_{13c} = +0.235 \pm 0.109$, while the sixth order coefficient is lower by only 0.094.

6. Taking all the results together they confirm, in general, but refine, those of the earlier study. It is now seen that a considerable part of the high correlation then found for the tuberculosis death rate with epidemicity index disappears if the cities are made constant and

equal in respect to some six important demographic and environmental factors. The correlation is still, however, in spite of the reduction, significant in comparison with its probable error. Death rate from all causes is highly correlated under all circumstances with explosiveness of outbreak. This death rate may be taken as an index of the general healthfulness of the community.[9] But the outstanding *single* factor, which apparently more than any other one thing yet discovered, determined how abruptly or explosively the mortality was to rise at the outbreak of the epidemic, was the normal death rate in the community from organic diseases of the heart. This extraordinary and striking fact will be more fully discussed in the next paper in this series.

VI. Summary.

In this second study a further more refined and detailed analysis is made of the weekly mortality statistics of the influenza epidemic of the autumn of 1918. Using the method of multiple or partial correlations the attempt is made to determine more critically the factors chiefly responsible for the great variation exhibited among 34 large American cities in respect to the degree of suddenness or explosiveness of outbreak of epidemic mortality. A new and more critically exact epidemicity index is described, as well as a more accurate single numerical index of the age distribution of a population. Every effort was made to get critical quantitative measures of the variables dealt with. In general, this more extended and refined study confirms the results of the first. It is believed that the critical refinements introduced in this and the succeeding studies fully meet the implied criticisms of Winslow and Rogers [10] on the first study, at least so far as concerns readers capable of understanding the implications of the theory of probability. It is, for example, simply idle to assert that a surely significant correlation between death rate from organic diseases of the heart and explosiveness of outbreak of influenza may be "due to peculiarities in age distribution of the population" when it is conclusively shown, as it here is, that the same correlation prevails *when the age distributions of all the populations concerned are held constant.* If the results of these studies are to be successfully controverted it must be upon a different basis than logic of this sort. Of the six environmental and demographic variables tested, only one, namely, the city's latitude, was found to have a statistically significant net correlation with the explosiveness or epidemicity index. In that case the biological or epidemiological significance of the result is doubtful, because of the peculiar non-

[9] It is so used, for example, by Brownlee, J., An Investigation into the Epidemiology of Phthisis in Great Britain and Ireland: Med. Res. Comm. Spec. Rept. Ser. No. 46, pp. 1-98, 1920.

[10] Winslow, C.-E. A., and Rogers, J. F., Statistics of the 1918 Epidemic of Influenza in Connecticut: Jour. Inf. Dis. vol. 26, pp. 184-216, 1920.

random spatial distribution of the 34 cities furnishing data. On the other hand, the normal death rate from all causes and from organic diseases of the heart are correlated to a relatively high degree with explosiveness of outbreak of epidemic mortality. The normal death rate from pulmonary tuberculosis is also sensibly, but not highly, correlated with explosiveness of influenza outbreak. The normal death rates from typhoid fever and from cancer show no sensible net correlation with explosiveness.

III. ON THE CORRELATION OF DESTRUCTIVENESS OF THE 1918 EPIDEMIC [11]

I. Introduction.

It was pointed out by friendly critics after the appearance of the first of these Studies [12] that the epidemiological character there dealt with was explosiveness of outbreak, as measured by an epidemicity index I_5 subsequently modified into I_6 as described in the second of these Studies [13], and that as explosiveness of outbreak of epidemic mortality was obviously a characteristic which might conceivably be nearly or quite distinct and independent from destructiveness as measured by the total number of persons killed by the epidemic, it was a matter of doubt what might be the real meaning of the biological conclusions flowing from the peculiar facts brought out in that first study, since confirmed and extended in the second. This criticism was well taken. The obvious answer to it, however, is to study destructiveness in the same way that we have explosiveness of outbreak, by the method of multiple correlation. It is the purpose of this paper to present the results of just such a study, in which the extended and critically defined series of variables of Influenza Studies II were used.

The problem to which this paper is addressed may then be stated in the following terms: The 34 large American cities for which we have available published weekly data, varied enormously among themselves in respect to the destructiveness of the epidemic, as measured by the number of their inhabitants who died in excess of the normal, during the autumn and winter of 1918–19. What factor significantly influenced or determined this variation?

The methods by which this problem is attacked in the present paper are precisely the same as those of the second of these Studies, and the reader is referred to the introductory portion of that paper for the general discussion of methods used.

---

[11] Papers from the Department of Biometry and Vital Statistics, School of Hygiene and Public Health, Johns Hopkins University. No. 22.

[12] Pearl, R., Influenza Studies. I. On Certain General Statistical Aspects of the 1918 Epidemic in American Cities: Public Health Reports, vol. 34, No. 32, pp. 1743-1783. 1919. Reprint No. 548, pp. 1-43. All citations in this paper are in terms of the pagination of the reprint.

[13] See p. 276.

II. Variables Discussed.

The variables discussed in the present paper are precisely the same as those listed in Influenza Studies II (p. 275) with two exceptions. These are (a) that here the variable I_6 (explosiveness of outbreak) is omitted, and (b) its place is taken by a new variable, destructiveness, which is indicated by the subscript 2 in all that follows. Destructiveness is measured by the 25-week excess mortality rates calculated and published by the Bureau of the Census.[14] These 25-week excess rates indicate the number of people dying from all causes, during the 25 weeks following the initial outbreak of the epidemic in this country in the autumn of 1918, in excess of the number who probably would have died in the same period had no epidemic occurred. The rates for the 34 cities are given in Table I (p. 12) of Influenza Studies I, and hence need not be reprinted here. The values of the other variables used in this paper are given in Table I of Influenza Studies II (p. 279).

III. Demographic and Environmental Correlations.

Just as in the discussion of explosiveness of outbreak, we may consider first the net fifth order correlations of destructiveness with the several demographic and environmental factors for which we have data. The coefficients are exhibited in Table I.

TABLE 1.—*Net correlation of destructiveness (25-week excess mortality) with various demographic and environmental factors.*

Variable correlated with destructiveness (25-week excess mortality).	r subscripts.	Coefficient.
Age distribution of population	24. 56789	+0. 132±0. 114
Sex ratio of population	25. 46789	+0. 161±0. 113
Density of population	26. 45789	+0. 163±0. 113
Latitude of city	27. 45689	−0. 424±0. 095
Longitude of city	28. 45679	−0. 133±0. 114
Rate of growth of population, 1900-1910	29. 45678	−0. 083±0. 115

The results here are clear cut. The only variable having a net coefficient which can be regarded as sensibly different from zero is latitude. There the coefficient is certainly statistically significant. Taken at its face value, and remembering the significance of the negative sign, this means that the farther south the city the greater the rate of mortality in excess of the normal. Or put in another way, the coefficient indicates that the influenza epidemic had a definite, though at best not marked, tendency to be more destructive of life in southern than in northern latitudes. Too much stress must not be laid upon this result, however, because of the peculiar spatial distribution of this group of 34 cities. This point has been fully discussed in the preceding study of this series and need not be

[14] *Cf.* Public Health Reports, vol. 34, No. 11, p. 505, 1919.

dwelt on again here. The same reasons exist for suspending judgment as to the full biological significance of the apparently definite correlation between destructiveness of the epidemic and latitude, as were emphasized in the case of the similar correlation between explosiveness of outbreak and latitude.

The remaining coefficients are very definitely so small as to leave no doubt about their meaning. Throughout the long process of making more and more variables constant and getting the successive higher order net coefficients, no one of these demographic characters, except latitude, ever showed a value sensibly differing from zero. This fact is well indicated by the zero order coefficients, which are as follows:

$$r_{24} = +0.024 \pm 0.116 \text{ (age)}.$$
$$r_{25} = -0.029 \pm 0.116 \text{ (sex)}.$$
$$r_{26} = +0.111 \pm 0.114 \text{ (density)}.$$
$$r_{27} = -0.325 \pm 0.103 \text{ (latitude)}.$$
$$r_{28} = +0.0007 \pm 0.116 \text{ (longitude)}.$$
$$r_{29} = -0.071 \pm 0.115 \text{ (growth)}.$$

It can, then, be safely asserted that in the determination of the variation among these 34 large American cities in respect to the excess mortality due to the epidemic, the age and sex distribution of the population, its degree of crowding (not its rate of recent growth), or the distance of the city west from the Atlantic seaboard, played no appreciable part whatever. This conclusion is true whether all these factors were allowed to vary together, or whether, as in a laboratory experiment, one to five of them were held constant, while the net influence of one or more varying alone was tested.

Whether or not the same conclusion will hold generally for other cities of a different order of size, or for rural districts, remains to be shown by further work. But as to the facts for the 34 cities listed in Table I of the second of these Studies, there can be no doubt or argument.

IV. Death Rate Correlations.

Turning to the same set of normal death rate variables as were studied in connection with explosiveness we have the results set forth in Table II.

TABLE II.—*Net correlation of destructiveness (25-week excess mortality) with the normal death rate from certain specified causes.*

Variable correlated with destructiveness (25-week excess mortality). Death rate from—	r subscripts.	Coefficient.
All causes	23f. 455789	+0.405±0.097
Pulmonary tuberculosis	23a. 455789	+0.279±0.107
Organic diseases of the heart	23b. 456789	+0.537±0.082
Nephritis and Bright's disease	23c. 455789	−0.008±0.116
Typhoid fever	23d. 455789	−0.133±0.113
Cancer and other malignant tumors	23e. 455789	+0.268±0.107

From this table we note the following points:

1. Epidemic excess mortality is significantly correlated with the normal death rate from all causes in these cities. This is true in gross ($r_{23f} = +0.435 \pm 0.094$) and also when the environmental and demographic factors listed in Table 1 above are held constant, as in an experiment. In other words, the number of people dying during the epidemic in each of these 34 cities was determined to a significant, though not a high, degree by the usual mortality relations of the community, as indicated by the normal death rate from all causes. Cities which have normally a high death rate had also a relatively high mortality from the influenza epidemic, and *vice versa* those normally enjoying a relatively low mortality rate lost but relatively few persons in the epidemic. It is to be noted, however, that while the net coefficient $r_{23f.456789}$ has a value more than 4 times its probable error and therefore is to be regarded as certainly significant statistically, yet this correlation is lower than the corresponding one for explosiveness of outbreak and death rate from all causes (cf. Table III, Influenza Studies II). We have:

Explosiveness, $r_{13f.456789}$	$= +0.572 \pm 0.078$
Destructiveness, $r_{23f.456789}$	$= +0.405 \pm 0.097$
Difference	0.167 ± 0.124

While this difference is not significant in comparison with its probable error, nevertheless there is considerable probability that with a larger sample it would become so. It thus appears that the explosiveness of outbreak of the epidemic mortality was, perhaps, somewhat more influenced by the normal mortality rate of the community than was its total magnitude or destructiveness.

2. The total destructiveness of the epidemic is not significantly correlated with the normal death rate of the community from pulmonary tuberculosis when the important demographic and environmental factors listed in Table 1 are held constant. Here we come upon a distinct break between the two epidemiological characteristics, explosiveness and total destructiveness. The former is, and the latter is not, significantly correlated with the normal tuberculosis death rate. A difference of the same sense is evident in the zero order gross correlations, which are $r_{13a} = +0.578 \pm 0.077$ and $r_{23a} = +0.428 \pm 0.094$. Both of these gross values are more than 3 times the probable error, but owe a considerable portion of their high values, as is demonstrated by the sixth order coefficients, to intercorrelations with other variables.

3. The highest net correlation of destructiveness of the epidemic is with the normal death rate from organic diseases of the heart. When all of the demographic and environmental factors with which

we are dealing are held constant for these cities, we find a correlation of $+0.537 \pm 0.082$, a value nearly 7 times its probable error, between these two variables. The coefficient is higher than that for the normal death rate from all causes with destructiveness. It appears very clearly that of the 12 different factors here studied the normal death rate of the community from organic diseases of the heart had more to do with determining the proportionate part of the population which perished in the epidemic than any other factor. In those cities having normally a high heart-disease rate a relatively large number died in the influenza epidemic, and *vice versa*. The same thing was shown in the preceding study to be true for explosiveness of outbreak of epidemic mortality. The condition of the population in respect to cardiac soundness played a significant rôle in determining the suddenness and frequency with which people died during the autumn and winter of 1918, when the 34 large American cities were struck by the influenza epidemic. It is interesting to note that the net sixth order correlation coefficient is higher in the case of organic diseases of the heart than the gross zero order coefficient, which is $r_{23b} = +0.487 \pm 0.088$. This means that in the gross, or zero order coefficient, the true organic relationship existing between destructiveness of epidemic and normal cardiac death rate is obscured by the fact that there is a high correlation between the latter variable and the age distribution of the population ($r_{3b_4} = +0.609 \pm 0.073$), the meaning of this coefficient being, of course, that the higher the average age of a population, the higher the death rate from organic diseases of the heart, and *vice versa*—a relationship which would be expected *a priori* from what we know about cardiac affections. As soon as we make the cities all constant in respect to age distribution of the population, we get a marked increase in the correlation coefficient between epidemic destructiveness and normal cardiac rate ($r_{23b.4} = +0.596 \pm 0.075$). This is an increase of 0.109 in the coefficient. But, as there is also a sensible negative correlation in this group of 34 cities between normal death rate from organic heart diseases and latitude ($r_{3b.7} = -0.300 \pm 0.105$), the high value of $r_{23b.4} = +0.596 \pm 0.075$ is reduced somewhat when the cities are made constant in respect to both age distribution and latitude, the coefficient being $r_{23b.47} = +0.549 \pm 0.081$. The other demographic and environmental variables have only negligible effect upon the r_{23b} correlation.

4. The net correlation between destructiveness of the epidemic and the normal death rate from diseases of the kidneys is sensibly zero. Here again there is a marked contrast between explosiveness of epidemic mortality and destructiveness (cf. Table III, Influenza Studies II). The gross zero order correlation between destructiveness and normal death rate from kidney diseases ($r_{23c} = +0.282 \pm$

0.106) is less than three times its probable error, as well as the net sixth order coefficient.

5. The destructiveness of the epidemic is not significantly correlated with the normal death rate from either typhoid fever or cancer, either net, when all the demographic and environmental variables are held constant, or in gross ($r_{23d} = +0.014 \pm 0.116$, and $r_{23e} = +0.215 \pm 0.110$). We observe the same contrast in the correlations of these diseases with destructiveness as we did in the previous study, when they were correlated with explosiveness, in comparison with the correlation of normal death rate from organic diseases of the heart with these same epidemiological characteristics.

V. Summary.

In order to meet a justifiable criticism of the earlier work we have determined in this paper the correlations between destructiveness of the influenza epidemic of 1918–19, as measured by the 25-week excess mortality in the 34 cities used in the earlier work, and the series of demographic, environmental, and biological (normal death rate) variables discussed in Studies II. It is found that there is, in the group of large cities, no significant net correlation between destructiveness of the epidemic as above defined, and any demographic or environmental variable, with the exception of latitude. In that case, for reasons fully discussed in Studies II, the biological meaning to be attached to the result is not entirely clear. The highest net correlation between destructiveness and any of the 12 variables discussed was with the normal death rate from organic diseases of the heart. Those cities, within this group of 34, which have normally an unusually high death rate from cardiac disorder, had an unusually high epidemic mortality, and *vice versa*. There was no significant net correlation between destructiveness of the epidemic and the normal death rate from acute nephritis and Bright's disease, nor with that from pulmonary tuberculosis. The correlation results indicate clearly that explosiveness of outbreak of epidemic mortality and total destructiveness of the epidemic are distinct but related epidemological characteristics.

IV. ON THE CORRELATION BETWEEN EXPLOSIVENESS AND TOTAL DESTRUCTIVENESS OF THE EPIDEMIC MORTALITY.[15]

The results of the second and third of these Studies indicate clearly that explosiveness of outbreak of epidemic mortality (as measured by the epidemicity index I_6) and total destructiveness of the epidemic (as measured by the 25-week excess mortality) are distinct epidemiological characters. The question then presents

[15] Papers from the Department of Biometry and Vital Statistics, School of Hygiene and Public Health, Johns Hopkins University. No. 23.

itself as to how closely these two characters are correlated. To the answer of this question the present paper is addressed.

For the 34 large American cities furnishing the material set forth in Table I of the second of these Studies,[16] we have data on the following variables:

Subscript No.	Variable.
1.	Explosiveness of outbreak of epidemic mortality, I_6.
2.	Destructiveness (25-week excess mortality).
3a.	Normal death rate from pulmonary tuberculosis.
3b.	Normal death rate from organic heart diseases.
3c.	Normal death rate from acute nephritis and Bright's disease.
3d.	Normal death rate from typhoid fever.
3e.	Normal death rate from cancer and other malignant tumors.
3f.	Normal death rate from all causes.
4.	Age distribution of population.
5.	Sex ratio of population.
6.	Density of population.
7.	Latitude.
8.	Longitude.
9.	Rate of growth of population, 1909-1910.

The gross or zero order correlation between explosiveness of epidemic mortality (I_6) and total destructiveness (25-week excess mortality) is

$$r_{12} = +0.709 \pm 0.058.$$

This obviously represents a relatively high, but by no means perfect, correlation. In general, it means that cities having an incidence of epidemic mortality more sudden and explosive in its outbreak than the average were highly likely to have a total mortality from the epidemic above the average, and *vice versa*.

It is essential, however, just as in earlier parts of these Studies, to find the *net* value of this correlation when the various environmental, demographic, and biological variables listed at the beginning of this paper are held constant as in an experiment. The initial step in such a process is to calculate the first-order coefficients, where successively, one at a time, each variable in the series is held constant while the correlation between variables 1 and 2, in which we are interested, is determined.[1] In Table I are given these first-order correlation coefficients between the variables 1 and 2 (explosiveness and destructiveness).

[16] See p. 279.

[1] For a general discussion of the method of multiple correlation in statistical work of this sort the reader is referred to these Studies, II, pp. 272-275.

TABLE 1.—*First-order correlations between explosiveness (I_6) and destructiveness (25-week excess mortality) of the 1918 influenza epidemic in 34 American cities.*

r subscripts.	Coefficient.
12.3a	+0.626±0.070
12.3b	+0.592±0.075
12.3c	+0.679±0.062
12.3d	+0.750±0.051
12.3e	+0.694±0.060
12.3f	+0.626±0.070
12.4	+0.718±0.056
12.5	+0.736±0.053
12.6	+0.707±0.058
12.7	+0.687±0.061
12.8	+0.729±0.054
12.9	+0.723±0.055

It is evident from this table that making any single one of the 12 variables constant has little effect upon the correlation between explosiveness and destructiveness of the epidemic mortality. The coefficients are all relatively high.

Let us examine the effect of making *all* the demographic and environmental factors (subscripts 4 to 9, inclusive) constant at the same time. The coefficient is

$$r_{12.456789} = +0.706 \pm 0.058.$$

This is almost absolutely identical with the zero order gross coefficient given above. This result means that the factors age and sex distribution, density and rate of recent growth of the population, latitude and longitude of the city (with all implied climatic differences), have no influence in determining the correlation between explosiveness and total excess mortality rate.

Taking in the biological (normal death rate) variables we have the following seventh-order coefficients:

$$r_{12.3a456789} = +0.675 \pm 0.063$$
$$r_{12.3b456789} = +0.579 \pm 0.077$$
$$r_{12.3c456789} = +0.711 \pm 0.057$$
$$r_{12.3d456789} = +0.749 \pm 0.051$$
$$r_{12.3e456789} = +0.700 \pm 0.059$$
$$r_{12.3f456789} = +0.632 \pm 0.069$$

It is obvious that these normal death rates influence very little, either one way or the other, the correlation between explosiveness and destructiveness of the epidemic outbreak.

One point of considerable interest which attaches to the relatively high correlation between these two variables is that one of the variables is measurable in time long before the value of the other can be possibly determined. The value of the explosiveness index I_6 can usually be determined in from 2 to 4 weeks after the beginning of the epidemic, while the total excess mortality can only be measured when the epidemic is all over. With a correlation coefficient of the

magnitude of r_{12} above, one can, by means of the regression equation, make a very fair prediction of the total excess mortality rate from a knowledge of the explosiveness of outbreak of the mortality measured

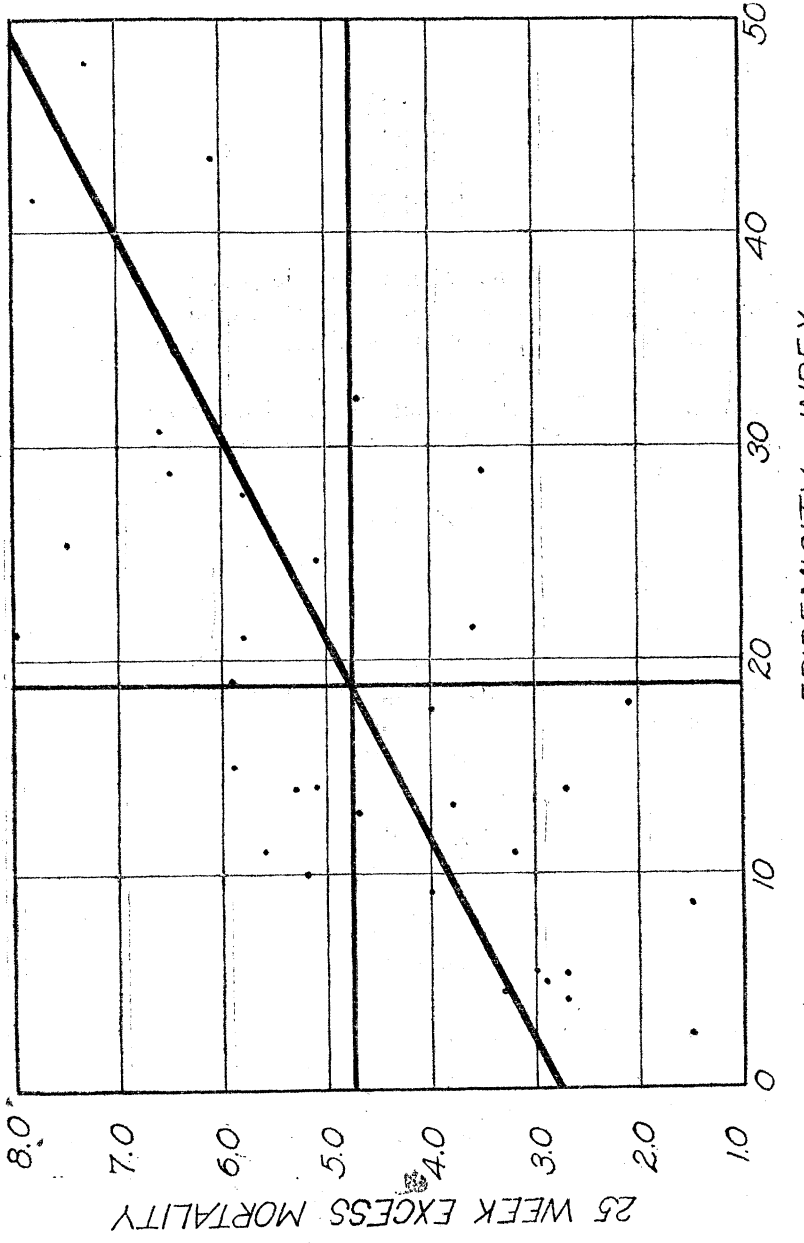

FIG. 2.—Regression line of 25-week excess mortality rate of influenza epidemic on explosiveness of outbreak of epidemic mortality.

by I_3. Actually, as the different net correlations show, we shall get practically as good a result by using the zero order coefficient and a simple equation of the form $y = a + bx$ as by employing a many-con-

stant partial regression equation. Actually the regression equation from the zero order r_{12} is·

$$D = 2.7412 + 0.1065\ I_6$$

where D denotes 25-week excess mortality rate and I_6 is the epidemicity or explosiveness index. This regression line is shown graphically in Figure 2.

SUGGESTED HEALTH PROVISIONS FOR STATE LAWS RELATING TO CHILDREN.

Report of the Advisory Committee to the National Child Health Council on Health Provisions for Laws Relating to Children.

In view of the fact that a number of States have children's code commissions at work, which are dealing with various aspects of child welfare, it is especially desirable to call attention at this time to the need for child-health provisions in order that they may receive the consideration that they merit. Therefore, there is presented here a report of the advisory committee to the National Child Health Council on Health Provisions for Laws Relating to Children.

The advisory committee is composed of the following members:

Courtenay Dinwiddie, chairman, executive secretary, National Child Health Council, Washington, D. C.

James A. Tobey, secretary, assistant director, department of health service, American Red Cross, Washington, D. C.

Richard A. Bolt, M. D., general director, American Child Hygiene Association, Baltimore, Md.

E. Dana Caulkins, manager, National Physical Education Service, Washington, D. C.

Taliaferro Clark, medical officer in charge of field investigations in child hygiene, United States Public Health Service, Washington, D. C.

Edward N. Clopper, assistant secretary, National Child Labor Committee, New York, N. Y.

Anna E. Rude, M. D., director, division of hygiene, Children's Bureau, United States Department of Labor, Washington, D. C.

Willard S. Small, Ph. D., specialist in school hygiene, Bureau of Education, Department of the Interior, Washington, D. C.

FOREWORD.

Inasmuch as health is of paramount importance to child life and as it has often received minor consideration in State children's codes, the Committee on Health Provisions for Laws Relating to Children wishes to emphasize the need for adequate treatment of this subject by all States. The following are points which should be borne in mind.